

Алгоритмическое обеспечение рекомендательных систем в сфере электронной коммерции

В. И. Пименов, email: v_pim@mail.ru¹

Т. С. Аникина, email: tanyanikina@gmail.com¹

И. В. Пименов, email: i-pim@mail.ru²

¹ Санкт-Петербургский государственный университет промышленных технологий и дизайна

² Государственный университет морского и речного флота имени адмирала С. О. Макарова

Аннотация. Рассматриваются алгоритмы создания персонализированных рекомендаций, в разной степени опирающиеся на данные о характеристиках товаров или услуг, оценки продуктов и профили пользователей. Предлагается подход к выявлению с помощью нейронной сети явных и латентных причинно-следственных связей между существенными характеристиками товара или услуги, его показателями качества и оптимизации рекомендаций на основе концепции доминирования.

Ключевые слова: электронная коммерция, маркетплейс, рекомендательная система, машинное обучение, нейронная сеть, коллаборативная фильтрация, кластеризация, Парето-оптимизация.

Введение

Электронная коммерция – активно развивающаяся сфера цифровой экономики, которая включает в себя финансовые и торговые транзакции, осуществляемые при помощи компьютерных сетей, и бизнес-процессы, связанные с проведением таких транзакций.

К электронной коммерции относят:

- электронную торговлю (e-trade),
- электронный маркетинг (e-marketing),
- электронный банкинг (e-banking),
- электронные страховые услуги (e-insurance).
- электронные деньги (e-cash) и т.д.

Площадки электронной торговли могут использовать разные схемы торговли: B2B (business-to-business) – “бизнес-бизнес”, B2C (business-to-consumer) – “бизнес-потребитель” и C2C (consumer-to-consumer) – “потребитель-потребитель”, представляющую торговые отношения между физическими лицами.

Самые распространенные предприятия электронной торговли – маркетплейсы, выступающие связующим звеном между потребителем и продавцом и работающие по схеме B2C.

Маркетплейс предлагает своим покупателям товары от разных продавцов, причем один и тот же товар может быть представлен несколькими продавцами, зарегистрированными на площадке. Наиболее известные маркетплейсы, относящиеся к сегменту B2C торговли – это российские интернет-компании OZON и WildBerries, китайский гипермаркет AliExpress, американские маркетплейсы Amazon и eBay.

Преимуществами модели электронной торговой площадки перед обычными интернет-магазинами являются широкий выбор товаров и удобство для покупателей.

Поскольку крупные маркетплейсы ежедневно обрабатывают сотни тысяч заказов, возникает проблема генерирования больших объемов данных и использования их скрытого потенциала для увеличения продаж. Ключевыми задачами при управлении сервисами торговой площадки становятся не просто создание аналитических отчетов и интерактивных визуализаций данных, но и выработка рекомендаций, основанных на методах принятия решений в условиях многокритериальности и многомерности данных.

Целью рекомендательных систем, использующих методы интеллектуального анализа данных, является расширение клиентской базы компании и повышение продаж за счет тщательно подобранных предложений.

Работы в сфере разработки рекомендательных систем связаны с применением глубокого машинного обучения и анализа временных рядов, позволяющих посредством рекомендаций предвидеть сезонные покупки, создавать персонализированные посадочные страницы для пользователей [1].

Примерно 35 % покупок на маркетплейсе Amazon являются результатом использования рекомендательной системы. Также около 75 % контента, который пользователи смотрят на стриминговом сервисе Netflix, основаны на рекомендациях [2].

1. Использование методов интеллектуального анализа данных на предприятиях электронной торговли

Методы интеллектуального анализа данных и алгоритмы машинного обучения могут применяться торговыми онлайн-площадками для решения различных задач, приведенных в таблице.

Одними из самых мощных систем искусственного интеллекта, которые предприятия электронной торговли внедряют для увеличения продаж, являются рекомендательные системы. Их основной целью

является выдача рекомендации по продуктам и услугам на основе предпочтений пользователя или его прошлых действий.

Таблица

Примеры применения интеллектуального анализа данных

Метод	Сфера, задача
Классификация	Веб-аналитика: прогнозирование оттока пользователей и обнаружение мошенничества. Торговля: построение модели предсказания рекомендательной системы. Финансовая сфера: оценка кредитоспособности
Регрессия	Финансовая сфера: оценка допустимого кредитного лимита
Прогнозирование временного ряда	Торговля: прогнозирование продаж. Финансовая сфера: прогнозирование цен на акции
Кластеризация	Торговля, финансовая сфера: сегментация клиентов, персонализация контента и построение рекомендательной системы
Аффинитивный анализ	Торговля: анализ потребительской корзины
Анализ последовательностей	Веб-аналитика: анализ переходов по страницам веб-сайта
Анализ отклонений	Веб-аналитика: обнаружение спама и мошенничества. Финансовая сфера: выявление мошенничества с банковскими картами

Для учета опыта взаимодействия пользователя и увеличения его вовлеченности с помощью рекомендательной системы требуется реализовать ряд этапов обработки данных, которые укрупненно выглядят следующим образом:

- извлечение данных из хранилища,
- преобразование данных в «плоский» файл надлежащего формата,
- выполнение сценария интеллектуального анализа, включающего совокупность методов Data Mining,
- интерпретация результатов анализа в терминах бизнес-концепций.

Построение рекомендательной системы – объемная и сложная задача, требующая современного алгоритмического обеспечения и больших вычислительных ресурсов. Для решения этого типа задач задействуется глубокое обучение нейронных сетей.

2. Построение прогнозов в рекомендательной системе

Рекомендательные системы нацелены на то, чтобы предсказать интересы пользователей и рекомендовать продукты, которые, скорее всего, будут им интересны.

Типичная архитектура рекомендательной системы приведена на рис. 1.

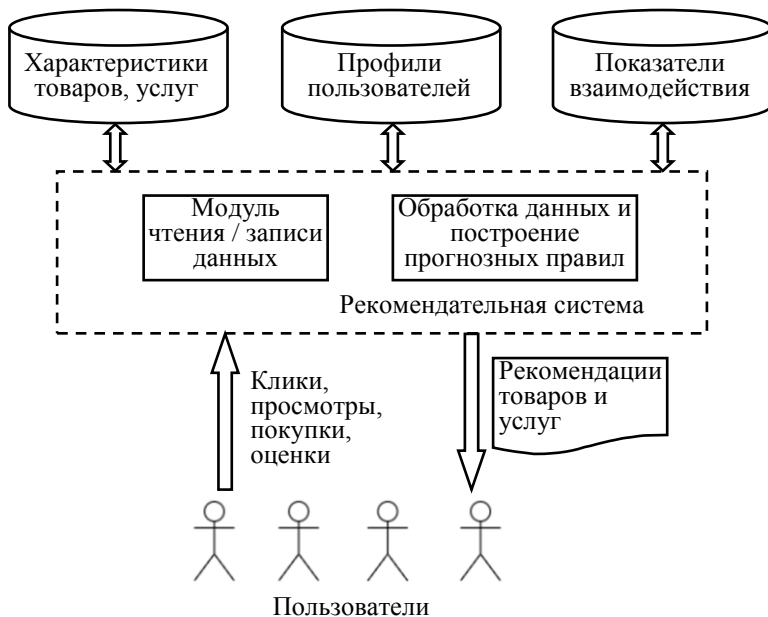


Рис. 1. Архитектура рекомендательной системы

Данные, используемые для построения рекомендаций, можно разделить на две категории:

- характеристики x товаров, услуг (функции, технические показатели, параметры, конструктивные особенности, цвет и материал, эргономичность, условия эксплуатации, условия предоставления услуги, цена) и профили пользователей u (возраст, пол, интересы, предпочтения),

- показатели качества y товаров или услуг, формируемые в процессе взаимодействия с ними пользователей (количество покупок, рейтинги, число лайков).

Показатели качества основаны на запросах поисковых систем и истории покупок, на явных оценках пользователей после просмотра фильма, прослушивания песни или на другой информации о взаимодействии пользователей и товаров. Такие сайты, как Spotify, YouTube или Netflix, используют такие данные, чтобы предлагать плейлисты, ежедневные миксы или давать рекомендации по видеоконтенту.

Каждый материальный или информационный продукт по показателю качества может быть классифицирован, например, как товар с низким, средним или высоким рейтингом. Соответственно, характеристики товаров или услуг могут использоваться как входные показатели при построении модели классификации, которая фиксирует взаимосвязь между особенностями продукта и значениями его показателя качества. Эту модель также можно использовать для прогнозирования того, как пользователи будут оценивать новые продукты.

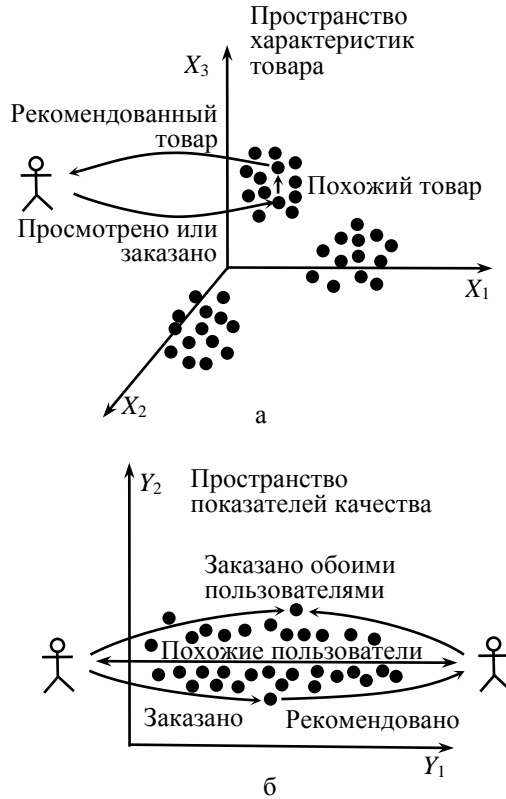
Рекомендательная система получает отзывы от конечных пользователей и запускает алгоритм генерации рекомендаций. Используя данные о пользователях, товарах и их взаимодействиях, алгоритм обновляет предпочтения пользователей. Затем система рекомендует список товаров или услуг, которые могут заинтересовать текущего пользователя, исходя из перечня его предпочтений.

Существуют различные алгоритмы создания персонализированных рекомендаций, в разной степени опирающиеся на данные определенной категории:

- алгоритмы фильтрации на основе контента, в которых используется профиль пользователя и характеристики товаров, услуг;
- алгоритмы совместной (коллаборативной) фильтрации, в которых ищется группа схожих пользователей и используются показатели, формируемые в процессе взаимодействия пользователей с предложениями торговой площадки;
- гибридные алгоритмы, объединяющие обе категории данных.

Алгоритмы фильтрации на основе контента дают рекомендации, используя профили пользователя и связанных товаров. Предполагается, что если пользователь интересовался товаром в прошлом, он заинтересуется аналогичным товаром в будущем (рис. 2, а). Похожие товары группируются на основе их характеристик, например, с

помощью алгоритмов кластеризации. Алгоритм может использовать новые товары без учета их рейтинга. Профили пользователей создаются при регистрации в системе путем явного опроса об их интересах, а также в ходе взаимодействий пользователей с системой или извлекаются из социальных сетей.



а – фильтрация на основе контента, б – совместная (коллаборативная) фильтрация

Рис. 2. Существующие подходы к фильтрации

Одна из проблем фильтрации на основе контента – выдача тривиальных рекомендаций из-за чрезмерной типизации. Например, если пользователя *А* интересуют только группы *В*, *С* и *Д*, то система

не может рекомендовать товары вне этих категорий, даже если они могут быть ему интересны. Еще одна распространенная проблема заключается в том, что новые пользователи не имеют определенного профиля, если только у них явно не запрошена информация о предпочтениях.

Совместная фильтрация в настоящее время является одним из наиболее часто используемых подходов и обычно дает лучшие результаты, чем рекомендации на основе контента. Такие алгоритмы для фильтрации интересующих пользователя товаров используют матрицу показателей качества $(y_{i,j})$, где каждый элемент $y_{i,j}$ представляет собой оценку i -м пользователем j -го товара. Если пользователь не оценил продукт, то соответствующий элемент остается пустым. На этапе формирования рекомендаций алгоритм рассчитывает для текущего пользователя вектор показателей качества, в котором пустые элементы заменяются прогнозными значениями на основе оценок, выставленных “похожими” пользователями.

Алгоритмы совместной фильтрации используют логические правила, например: если пользователю нравится товар A , а другому пользователю нравится тот же товар A и товар B , то первый пользователь также может быть заинтересован в товаре B (рис. 2, б). Таким образом, система предсказывает новые предпочтения на основе истории взаимодействий различных пользователей.

Ключевым шагом совместной фильтрации является выбор похожего пользователя. Ближайшие пользователи определяются с использованием той или иной меры сходства, которая рассчитывается с помощью матрицы $(y_{i,j})$.

Поскольку для новых пользователей отсутствует информация о взаимодействиях, а также большинство пользователей не ставят оценки товарам, то основной проблемой систем, применяющих совместную фильтрацию, является проблема холодного старта.

3. Подход, основанный на модели преобразования пространства характеристик в пространство показателей качества

Предлагаемый авторами подход нацелен на выявление явных и латентных причинно-следственных связей между существенными характеристиками товара или услуги $\{X_1, \dots, X_N\}$ и показателями качества продукта $\{Y_1, \dots, Y_K\}$.

Построение модели преобразования $x \rightarrow y$ (рис. 3) является трудноформализуемой задачей по ряду признаков: принципиальной невозможности точного описания всех взаимосвязей (часто

нелинейных) между характеристиками продукта и показателями качества, большой размерности пространства $x \times y$, динамически изменяющимся информационным потоком на разных стадиях жизненного цикла рекомендательной системы, ограниченным финансовым ресурсам [3].

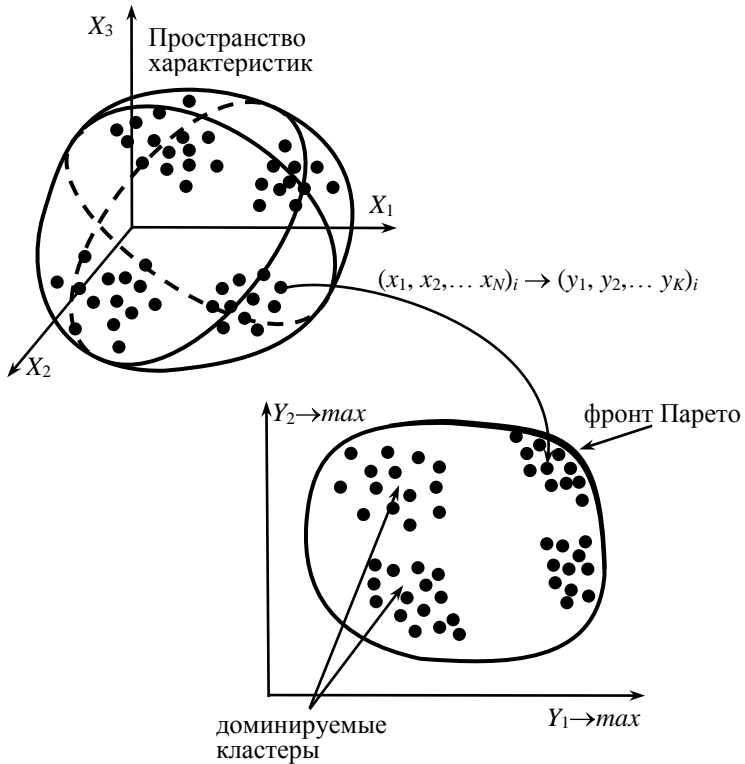


Рис. 3. Преобразование пространства характеристик в пространство показателей качества $x \rightarrow y$ ($N = 3, K = 2$)

Построение модели взаимосвязей $x \rightarrow y$ может быть осуществлено на основе методов машинного обучения, в рамках которых решаются задачи автоматической классификации продуктов, и упорядочения групп товаров по уровню качества.

Каждый продукт представляется в двух векторных пространствах – пространстве характеристик x и пространстве показателей качества y . После применения автоматической классификации, которая основана на гипотезе компактности, распределение близких товаров по однородным группам должно сохранять в пространствах x и y топологическое подобие. При оптимизации показателей качества поиск осуществляется в пространстве характеристик таким образом, что каждый вариант может быть прослежен в пространстве оценок и оценен на основе концепции доминирования [4]. Множество наилучших предложений в пространстве оценок чаще всего соответствует определенному кластеру в пространстве характеристик.

Автоматическая классификация в пространстве $x \times y$ требует наличия данных значительного объема, отсутствующих, как правило, в связи с холодным стартом. Обогащение выборки с сохранением явной и неявной связи между характеристиками товара и его оценками качества может выполняться с помощью прогнозирующей нейросетевой модели, обученной на наборе доступных прецедентов.

Для уплотнения ядра Парето-оптимальных рекомендуемых товаров, представляющих в многомерном пространстве y фронт Парето, используется набор генетических алгоритмов [5].

Заключение

Многомерность материального или информационного продукта включает агрегированное описание его характеристик, а также показатели качества. Данные, накопленные в базе данных торговой площадки, позволяют установить взаимосвязь показателей качества и выработать решения о рекомендациях товаров или услуг. При многокритериальной оптимизации поиск осуществляется в пространстве оценок продукта таким образом, что каждый вариант может быть прослежен в пространстве y и оценен на основе концепции доминирования.

Список литературы

1. At Alibaba, Artificial Intelligence is Changing How People Shop Online [Электронный ресурс]: Alizila. News from Alibaba. – Режим доступа: <https://www.alizila.com/at-alibaba-artificial-intelligence-is-changing-how-people-shop-online>
2. How retailers can keep up with consumers [Электронный ресурс]: McKinsey&Company. – Режим доступа: <https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers>

3. Пименов, В. И. Анализ и визуализация данных в задачах многокритериальной оптимизации проектных решений / В. И. Пименов, И. В. Пименов // Информатика и автоматизация (Труды СПИИРАН). – 2022. – Т. 21. – № 3. – С. 543-571.

4. Ayaz, Ahsan. Multi-Objective Design Optimization of Multi-copter using Genetic Algorithm / Ahsan Ayaz, Ashhad Rasheed // IEEE International Bhurban Conference on Applied Sciences and Technologies (IBCAST). – 2021. – Access. no. 20633290.

5. Pimenov, V. I. Interpretation of a trained neural network based on genetic algorithms / V. I. Pimenov, I. V. Pimenov // Информационно-управляющие системы. – 2020. – № 6. – С. 12-20.